

استفاده از آنتروپی شانون در پیش‌پردازش ورودی شبکه بیزین جهت مدل‌سازی سری‌های زمانی

فاطمه آخونی پورحسینی^۱، محمدعلی قربانی^۲ و کاکاشاهی^۳

۱- دانشجوی ارشد مهندسی منابع آب، دانشگاه تبریز، (نویسنده مسوول: fateme.pourhosseini@yahoo.com)

۲- دانشیار، مهندسی منابع آب، دانشگاه تبریز

۳- دانشیار، دانشکده منابع طبیعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری

تاریخ دریافت: ۹۵/۳/۱۱ تاریخ پذیرش: ۹۵/۱۱/۶

چکیده

انتخاب ورودی‌های مناسب برای مدل‌های هوشمند از اهمیت بسزایی برخوردار است. زیرا باعث کاهش هزینه و صرفه‌جویی در وقت و افزایش دقت و کارایی مدل‌ها می‌شود. هدف از این مطالعه، کاربرد آنتروپی شانون برای انتخاب ترکیب بهینه متغیرهای ورودی در مدل‌سازی سری زمانی می‌باشد. سری زمانی ماهانه بارش، دما و تابش در دوره زمانی ۱۳۶۱ تا ۱۳۸۹ برای ایستگاه سینوپتیک تبریز مورد استفاده قرار گرفت. پارامترهای بارش، دما و تابش با تأخیرهای مختلف به‌عنوان ورودی به آنتروپی شانون در نظر گرفته شد. نتایج آنتروپی شانون نشان داد که سری زمانی با سه تأخیر، نتایج بهتری را برای مدل‌سازی ارائه می‌دهد. شبیه‌سازی با استفاده از دو مدل شبکه‌ی بیزین و رگرسیون خطی چند متغیره انجام گرفت. کارایی مدل‌ها با استفاده از سه معیار: ضریب تبیین (R^2)، ریشه جذر میانگین خطا (RMSE) و شاخص پراکندگی (SI) محاسبه گردید. از میان این دو مدل با ساختار ورودی‌های یکسان، مدل شبکه عصبی بیزین عملکرد بهتری برای شبیه‌سازی سری زمانی بارش، دما و تابش در مقایسه با رگرسیون چندمتغیره داشته است. نتایج تحقیق نشان داد که آنتروپی شانون در انتخاب ترکیب ورودی مناسب برای مدل‌های هوشمند می‌تواند کارایی بهتری داشته باشد.

واژه‌های کلیدی: آنتروپی، رگرسیون خطی چندمتغیره، سری زمانی، شبکه بیزین

مقدمه

در تعیین بهینه نمونه‌برداری (۱۱)، طراحی و ارزیابی شبکه‌های پیمایش کیفیت آب (۱۲، ۱۰، ۱۳)، طراحی و ارزیابی شبکه ایستگاه‌های هیدرومتری (۱۶، ۳)، پیش‌بینی کیفیت آب رودخانه (۱۲)، تخمین پارامترهای توزیع فراوانی (۲۲، ۲۱)، تفکیک مقادیر بارش روزانه به مقیاس ساعتی (۱۸)، بررسی و ارزیابی خصوصیات زمانی و مکانی بارش (۱۵، ۲) استفاده شده است. امروزه و اسپیلدورا (۱) آنتروپی را برای ارزیابی عدم قطعیت مدل حوضه استانفورد^۳ در شبیه‌سازی رواناب یک حوضه در کالیفرنیا به کار بردند. سوناگا و همکاران (۲۲) اصل حداکثر آنتروپی را برای مدل‌سازی فرآیند بارش و رواناب به کار بردند. ایشان عبارت آنتروپی شرطی را بر اساس قیود معین برای به دست آوردن توزیع شرطی رواناب بر اساس رخداد بارش معین پیشنهاد نمود، آنگاه پارامترهای تابع توزیع مذکور را بر اساس اطلاعات موجود شامل میانگین سری زمانی داده‌های بارش و رواناب و کوواریانس آن‌ها تخمین زد. کریمی (۱۴) در حوضه آبریز زاینده‌رود، پس از محاسبه آنتروپی انتقال اطلاعات زوج ایستگاه‌ها، رابطه آنتروپی انتقال اطلاعات فاصله برای منطقه را تعیین و سپس با دو هدف پیشینه کردن حداقل آنتروپی و پیشینه کردن متوسط آنتروپی، مناطق نیازمند ایستگاه‌های جدید را معرفی کردند. چیانگ و همکاران (۶) تئوری آنتروپی انتقال اطلاعات در طراحی شبکه باران‌سنجی مورد استفاده قرار دادند. مطالعه‌ی موجود در مورد اثر تغییرات مکانی و زمانی بارش با محاسبه حداکثر آنتروپی مشترک می‌باشد. نتایج بیانگر این است که اثر تغییرات مکانی و مقایسه با مقیاس‌های زمانی بارش با محاسبه حداکثر آنتروپی مشترک می‌باشد. نتایج بیانگر این است که اثر تغییرات مکانی و مقایسه با مقیاس‌های زمانی ناچیز است.

امروزه سیستم‌های هوشمند برای پیش‌بینی پدیده‌های غیرخطی استفاده می‌شود، که روش شبکه بیزین (BN)^۱ و شبکه عصبی مصنوعی (ANN)^۲ از جمله‌ی این روش‌هاست. در سال‌های اخیر مدل‌های غیرخطی مانند شبکه بیزین (BN) به‌طور گسترده‌ای برای مدل‌سازی سری زمانی استفاده گردیده‌اند. فرج‌زاده و همکاران (۹) با استفاده از مدل‌های سری زمانی و شبکه عصبی مصنوعی به پیش‌بینی جریان ماهانه و بارش حوضه دریاچه ارومیه پرداختند. نتایج حاصل از پیش‌بینی جریان حاکی از دقت خوب هر دو مدل در تخمین جریان ماهانه دارد. البته طبق این تحقیق تفاوت چندانی بین دو مدل مقایسه در تخمین جریان وجود ندارد. کارمونا و همکاران (۵) برای مدیریت منابع آب‌های زیرزمینی در دو منطقه موردی واقع در اسپانیا از الگوی شبکه بیزین به‌عنوان یک سامانه مناسب پشتیبانی استفاده نموده‌اند. باتسیس و همکاران (۳) اثر تبخیر و تعرق را بر بارش- رواناب و ارتباط بین آن‌ها را با استفاده از آمار روزانه در حوضه آبخیز شمال کالیفرنیا مطالعه کردند و نشان دادند که مدل شبکه بیزین برای مطالعه وابستگی بارش و رواناب مناسب بوده و نیز تخمینی مؤثر برای نتایج میزان جریان بر اساس ورودی‌های مختلف می‌باشد. فرمانی و همکاران (۸) با شبکه بیزین به مدیریت بهینه آلودگی آب‌های زیرزمینی در کهپاک پرداختند. نتیجه به‌دست‌آمده حاکی از موفقیت‌آمیز بودن کاربرد الگوهای شبکه بیزین در مدیریت آلودگی آب‌های زیرزمینی منطقه بوده است. تئوری آنتروپی اولین بار توسط شانون توسعه داده شد و بعد از آن به‌طور گسترده در مسائل مختلف علمی مورد استفاده قرار گرفت. در زمینه‌ی هیدرولوژیکی از مفهوم آنتروپی

1- Bayesian Network

2- Artificial Neural Network

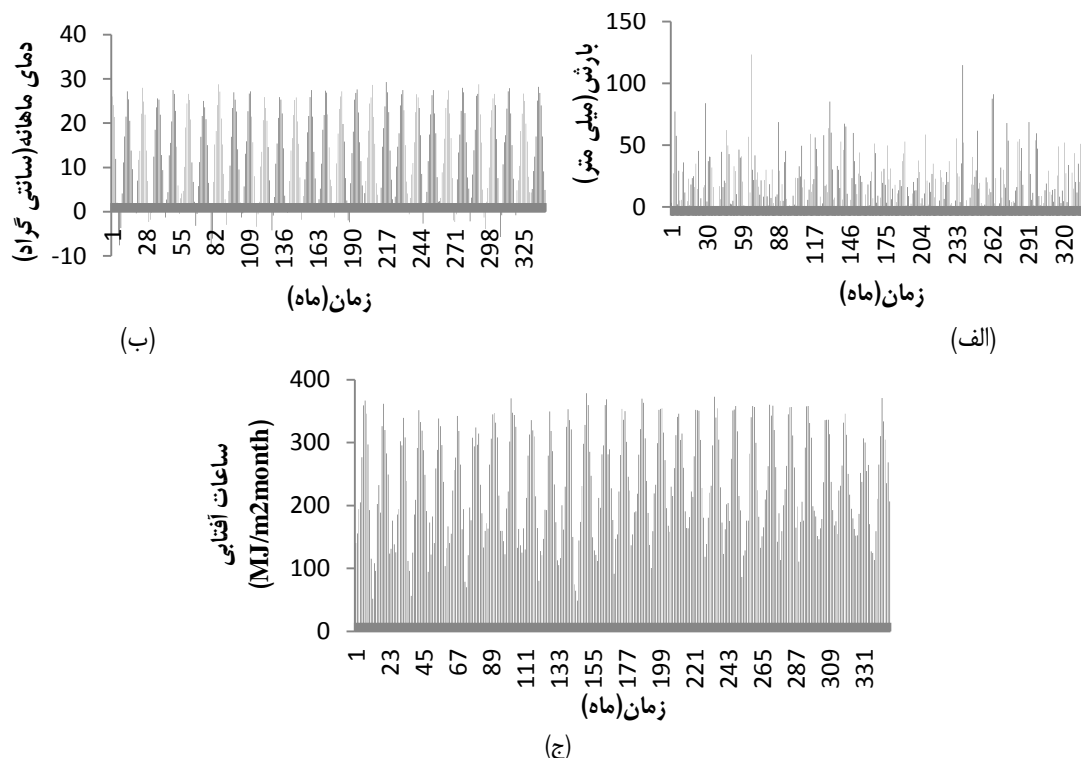
3- Stanford

مواد و روش‌ها

منطقه مورد مطالعه و داده‌های مورد استفاده

در این پژوهش، ایستگاه هواشناسی سینوپتیک تبریز به‌عنوان منطقه‌ی مطالعاتی انتخاب شد. که دارای طول جغرافیایی، ۴۶ درجه و ۱۷ دقیقه شرقی و عرض جغرافیایی ۳۸ درجه و ۵ دقیقه شمالی و ۱۳۶۱ متر ارتفاع می‌باشد. از داده‌های ماهانه‌ی ایستگاه سینوپتیک تبریز شامل بارش، دما و تابش طی دوره آماری (۱۳۸۹-۱۳۶۱) استفاده گردید که از مرکز آمار و اطلاعات هواشناسی اخذ گردید. در شکل (۱) نمودار بارش، دما و تابش و در جدول (۱) خصوصیات آماری پارامترهای مذکور در بازه‌ی زمانی ماهانه نشان داده شد.

رمضان و همکاران (۱۹) در حوضه برو^۱ انگلستان به‌منظور پیش‌بینی بارش و رواناب از آن‌تروپی شانون و آزمون گاما جهت تعیین ورودی‌ها به شبکه عصبی مصنوعی استفاده کردند که نتایج از عملکرد بهتر آن‌تروپی شانون برای تعیین ورودی‌ها حاکی بود. چن (۷) با استفاده از ماشین بردار پشتیبان و تئوری آن‌تروپی انتقال اطلاعات به استخراج داده‌های هیدرولوژیکی از حوادث سیل در حوضه رودخانه یانگ در تایوان پرداختند. در این تحقیق برای یافتن بهترین ترکیب از متغیرهای ورودی در پیش‌بینی سری زمانی بارش، دما و تابش در ایستگاه سینوپتیک تبریز از تئوری آن‌تروپی شانون استفاده شد و برای مدل‌سازی از شبکه عصبی بیزین و رگرسیون خطی چند متغیره استفاده گردید.



شکل ۱- مقادیر مشاهداتی بارش، دما و تابش ماهانه ایستگاه سینوپتیک تبریز
Figure 1. The observed values of monthly precipitation, temperature and radiation from Tabriz synoptic station

جدول ۱- مشخصات آماری بارش، دما و تابش ماهانه ایستگاه سینوپتیک تبریز

Table 1. Statistical Profile monthly precipitation, temperature and radiation from Tabriz synoptic station

پارامتر	واحد	تعداد داده	میانگین	مینیمم	ماکزیمم	انحراف معیار	چولگی	ضریب تغییرات
P _t	mm	۳۴۲	۲۰/۴۸	۰	۱۲۳/۴	۲۰/۷۶	۱/۵	۱/۰۱
T _t	C	۳۴۲	۱۳/۱۱	-۷/۹	۲۹/۳	۹/۸۹	-۰/۰۹	-۰/۷۵
S _t	MJ/m ² month	۳۴۲	۲۳۲/۹۴	۴۸/۹	۳۷۸/۶	۸۲/۹۲	۰/۰۱	۰/۳۶

شش ماه قبل به عنوان ورودی‌های مناسب به آن‌تروپی شانون معرفی گردید.

به‌منظور تخمین بارش، دما و تابش ماه‌های آینده، از اطلاعات ماه‌های قبل استفاده گردید. با استفاده از ماتریس همبستگی که در جدول (۲) مشاهده می‌شود تأخیر زمانی تا

جدول ۲- همبستگی متقابل بین پارامترهای ورودی و خروجی

Table 2. Cross-correlation coefficient between input and output parameters

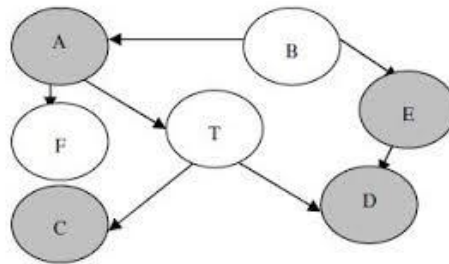
P_{t-6}	P_{t-5}	P_{t-4}	P_{t-3}	P_{t-2}	P_{t-1}	P_t
-۰/۱۱	-۰/۱۲	-۰/۱۵	-۰/۱۶	۰/۱۱	۰/۲۷	P_t
T_{t-6}	T_{t-5}	T_{t-4}	T_{t-3}	T_{t-2}	T_{t-1}	T_t
-۰/۱۵	-۰/۴۵	-۰/۴۷	۰/۳۳	۰/۴۸	۰/۸۴	T_t
S_{t-6}	S_{t-5}	S_{t-4}	S_{t-3}	S_{t-2}	S_{t-1}	S_t
-۰/۲۲	-۰/۳۲	-۰/۴۲	۰/۲۲	۰/۴۶	۰/۸۷	S_t

شبکه بیزین (BN)^۱

شبکه بیزین یک الگوی نموداری احتمالاتی است که مجموعه‌ای از متغیرها و احتمالات مربوط به هریک را نشان می‌دهد. یک گراف مستقیم و چرخه‌ای است که در آن، گره‌ها در حکم متغیرهای مسئله هستند. ساختار یک شبکه بیزین در واقع یک نمایش نموداری از اثرات متقابل متغیرهایی است که باید الگوبندی شوند و علاوه بر اینکه کیفیت رابطه بین متغیرهای مسئله را نشان می‌دهد، کمیت ارتباط بین این متغیرها را نیز به نمایش می‌گذارد که به صورت عددی از توزیع احتمال مشترک آن‌ها استفاده می‌کند. این روش بر مبنای محاسبات احتمالات شرطی (قانون بیز) بوده که معادله زیر رابطه بیز را نشان می‌دهد.

$$P(a|b) = \frac{P(b|a) * P(a)}{P(b)} \quad (1)$$

که در آن $P(a)$ احتمال وقوع پیشامد a ، $P(b)$ احتمال وقوع پیشامد b ، $p(a|b)$ احتمال شرطی پیشامد a است به شرطی که پیشامد b اتفاق افتاده باشد و $p(b|a)$ احتمال شرطی پیشامد b است به شرطی که پیشامد a اتفاق افتاده باشد. هر شبکه بیزین از سه جزء اصلی تشکیل شده است: مجموعه‌ای از گره‌ها، مجموعه‌ای از بندها و مجموعه‌ای از احتمالات، شکل ۲ نمونه‌ای از گراف را به صورت چرخه‌ای و مستقیم نشان می‌دهد.



شکل ۲- نمایش گراف مستقیم و چرخه‌ای شبکه بیزین
Figure2. The view of direct and cyclic Bayesian network

سیستم خبره عمل نماید. در این حالت در صورت عدم وجود داده‌های مشاهداتی کافی، تحلیل‌گر می‌تواند احتمالات شرطی بین علت و معلول را تعیین نموده و مدل را تحلیل کند. البته واضح است که اگر مدل شبکه بیزین بر پایه اطلاعات مشاهداتی بسیار اندک توسعه داده‌شده و تحلیل شود، خطای تحلیل و پیش‌بینی‌های مربوطه افزایش می‌یابد.

رگرسیون خطی چند متغیره

تحلیل رگرسیون روش آماری است که در آن از رابطه بین دو یا چند متغیر کمی (متغیرهای مستقل و یا پیش‌بینی کننده) برای پیش‌بینی متغیر وابسته (متغیر پاسخ) استفاده می‌شود. یک مدل رگرسیونی خطی چند متغیره به صورت زیر بیان می‌شود:

$$y = \sum_{i=1}^N \beta_i X_i + \varepsilon \quad (2)$$

که در آن β_i ضریب رگرسیون (پارامترهای رگرسیون)، X_i متغیر مستقل، ε خروج از مرکز و N تعداد متغیرهای مستقل است. به منظور تخمین ضرایب رگرسیون (β_i) از روش حداقل مربعات خطا استفاده می‌شود.

در حالت کلی گره‌ها یا والد هستند یا ولد (فرزند) یک گره فرزند می‌تواند توسط چندین والد، تولیدشده باشد گره‌هایی که قبل از آن‌ها گره دیگری در گراف وجود دارد با توزیع احتمال شرطی تعریف می‌شوند. در غیر این صورت با احتمال آغازین (اولیه) آن‌ها بیان می‌شوند. احتمالات مربوط به پایین‌ترین بخش در شبکه بیزین از طریق قانون احتمال کل به دست می‌آید و احتمالات مرتبط با بخش‌هایی بالایی شبکه، بر اساس قانون بیزین می‌باشد.

همان‌گونه که در بالا اشاره شد شبکه‌های بیزین امکان محاسبات روبه‌جلو و رو به عقب را برای تحلیل‌گر ایجاد می‌کنند. در واقع نه تنها از تجمیع وضعیت پارامترهای علت می‌توان به وضعیت معلول رسید بلکه در این روش با در اختیار داشتن وضعیت معلول یا همان پارامتر پیش‌بینی شونده با یک فرآیند بازگشت به عقب، امکان محاسبه وضعیت پارامترهای تأثیرگذار وجود داشته و به بیان دیگر می‌توان تعیین نمود که میزان تأثیرگذاری هر پارامتر بر ریسک نهایی یا همان خروجی نهایی چه قدر خواهد بود. علاوه بر موارد ذکرشده مدل شبکه‌های بیزین در بسیاری از موارد می‌توان به شکل

به دو قسمت تقسیم‌بندی شدند. ۷۰ درصد برای آموزش مدل و ۳۰ درصد برای بخش صحت سنجی استفاده گردید. به‌منظور افزایش کارایی مدل، همه داده‌ها ابتدا به شکل نرمال بین دو عدد صفر و یک استاندارد شدند سپس آن‌ها به مقادیر اولیه بعد از استفاده در شبیه‌سازی برگشتند.

$$X_n = \frac{X_i - X_{max}}{X_{min} - X_{max}} \quad (۸)$$

در این X_i رابطه داده‌های مشاهداتی، X_n داده‌های استاندارد شده، X_{min} و X_{max} ترتیب معرف داده‌های حداکثر و حداقل می‌باشد.

برای مقایسه داده‌های مشاهداتی و پیش‌بینی‌شده از شاخص‌های آماری ضریب تبیین (R^2)، ریشه متوسط مربعات خطا (RMSE) و شاخص پراکندگی (SI) استفاده گردید. به‌منظور مقایسه نتایج ماشین بردار پشتیبان و شبکه عصبی مصنوعی از شاخص‌های آماری نظیر: ضریب تبیین (R^2)، ریشه متوسط مربعات خطا (RMSE) و شاخص پراکندگی (SI) استفاده شد.

$$R^2 = \left[\frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \right]^2 \quad (۹)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad (۱۰)$$

$$SI = \frac{\frac{1}{n} \sum_{i=1}^n |x_i - y_i|}{\bar{x}} \quad (۱۱)$$

که در روابط ۹ الی ۱۱، X_i مقدار دبی مشاهداتی و y_i مقدار دبی پیش‌بینی‌شده و \bar{x} و \bar{y} به ترتیب نشانگر مقادیر میانگین داده‌های متناظر می‌باشند.

نتایج و بحث

به‌منظور تعیین ترکیب بهینه متغیرهای ورودی به شبکه بیزین و رگرسیون خطی چندمتغیره از آنروپی شانون استفاده گردید. با استفاده از آنروپی شانون مقدار آنروپی هریک از پارامترهای ورودی تعیین شد. سپس با توجه به این که مقدار بیشتر آنروپی بیان‌گر تأثیرگذاری بیشتر این پارامتر نسبت به سایر پارامترها در مدل‌سازی می‌باشد. بنابراین با توجه به شکل ۳ نتیجه می‌شود که تأثیر $P_{(t-1)}$ ، $P_{(t-2)}$ و $P_{(t-3)}$ در مدل‌سازی بارش و تأثیر $T_{(t-1)}$ ، $T_{(t-2)}$ و $T_{(t-3)}$ در مدل‌سازی دما و تأثیر $S_{(t-1)}$ ، $S_{(t-2)}$ و $S_{(t-3)}$ در مدل‌سازی تابش بر میزان خروجی بیشتر بوده و می‌توان به‌عنوان پارامترهای ورودی به مدل‌ها معرفی گردد که شکل ۳ هم این گفته تأیید می‌کند. جدول شماره ۳ و ۴ مقادیر شاخص آماری محاسبه‌شده مربوط به نتایج بخش آموزش و تست در شبکه بیزین و رگرسیون خطی چندمتغیره نشان می‌دهد.

تئوری آنروپی شانون

شانون (۲۰) نشان داد که وقایع با احتمال وقوع زیاد اطلاعات کمتری در اختیار می‌گذارند و برعکس هرچقدر احتمال وقوع یک رخداد کمتر باشد، اطلاعات حاصل از آن بیشتر است. با به دست آوردن اطلاعات جدید، درواقع عدم قطعیت‌ها کاهش‌یافته و ارزش اطلاعات جدید برابر با مقداری است که از عدم قطعیت کاسته شده است. درنتیجه عدم قطعیت و اطلاعات پارامترهایی وابسته به هم هستند. به‌عبارت‌دیگر، از تئوری آنروپی می‌توان به‌عنوان شاخصی برای کمی کردن میزان عدم آگاهی و دانش نسبت به مشخصات یک سامانه، استفاده کرد (۱۷).

برای به دست آوردن وزن آنروپی ω_j از معادله زیر استفاده می‌شود:

$$\omega_j = \frac{1 - e_j}{\sum_{j=1}^n 1 - e_j} \quad (۳)$$

پارامتر e_j مقدار آنروپی انتقال اطلاعات را بیان می‌کند درواقع آنروپی انتقال اطلاعات نوعی از آنروپی است که برابر با تفاضل مجموع آنروپی مرزی و آنروپی‌های مشترک می‌باشد. هرچقدر مقدار آنروپی کمتر باشد، تأثیر j بیشتر خواهد بود. آنروپی انتقال اطلاعات بین دو متغیره i و j به‌صورت زیر تعریف می‌شود:

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^m P_{ij} \ln P_{ij} \quad (۴)$$

با توجه به اینکه پارامترهای ورودی دارای واحدهای یکسانی باشند معمولاً با استفاده از تابع نرمال‌سازی نرمال‌سازی به‌صورت رابطه (۵) خواهد بود:

$$y_{ij} = \frac{x_{ij} - (x_{ij})_{\min}}{(x_{ij})_{\max} - (x_{ij})_{\min}} \quad (۵)$$

ماتریس داده‌های خام پس از نرمال‌سازی بدین‌صورت خواهد شد:

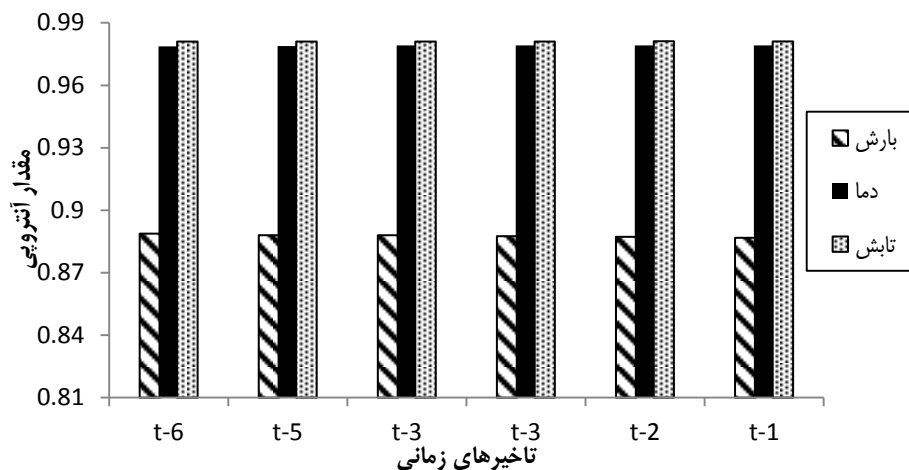
$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \dots & \dots & \dots & \dots \\ y_{m1} & y_{m2} & \dots & y_{mn} \end{bmatrix} \quad (۶)$$

نسبت مقدار شاخص پارامتر j در نمونه i بر اساس معادله زیر محاسبه‌شده است:

$$P_{ij} = y_{ij} / \sum_{i=1}^m y_{ij} \quad (۷)$$

تجزیه و تحلیل داده‌ها

برای استفاده از داده‌های مذکور در مدل‌های شبکه بیزین و رگرسیون خطی چند متغیره ابتدا داده‌ها به‌صورت تصادفی



شکل ۳- مقدار آنتروپی هر یک از پارامترهای ورودی به مدل
Figure 3. The entropy value of each parameter input to the model

جدول ۳- نتایج شبکه بیزین در مراحل آموزش و صحت‌سنجی

Table 3. Results Bayesian network used in the training and validation

تست			آموزش			پارامترهای ورودی	ترکیب	پارامتر
RMSE(mm, °C, MJ/m ² month)	SI	R ²	RMSE(mm, °C, MJ/m ² month)	SI	R ²			
۱۹/۴۹	-/۲	-/۵۵	۲-/۱	-/۱۸	-/۶۰	P _{t-1}	۱	P
۱۹/۵۷	-/۱۹	-/۶۰	۱۹/۸	-/۱۷	-/۶۳	P _{t-1} , P _{t-2}	۲	
۱۵/۶۳	-/۱۸	-/۸۵	۱۹/۷	-/۱۵	-/۸۰	P _{t-1} , P _{t-2} , P _{t-3}	۳	
۲/۸۵	-/۳۹	-/۷۱	۵/۲۴	-/۴۰	-/۷۲	T _{t-1}	۱	T
۲/۷۱	-/۲۴	-/۹۱	۲/۹۰	-/۲۴	-/۹۱	T _{t-1} , T _{t-2}	۲	
۲/۶۷	-/۱۶	-/۹۳	۲/۳۰	-/۱۶	-/۹۳	T _{t-1} , T _{t-2} , T _{t-3}	۳	
۳۶/۶۱	-/۲۴	-/۶۲	۵۲/۹۵	-/۲۵	-/۶۱	S _{t-1}	۱	S
۲۲/۶	-/۲۳	-/۶۱	۳۷/۸۴	-/۲۳	-/۶۷	S _{t-1} , S _{t-2}	۲	
۱۷/۴۳	-/۱۹	-/۷۸	۴۲/۳	-/۱۷	-/۷۵	S _{t-1} , S _{t-2} , S _{t-3}	۳	

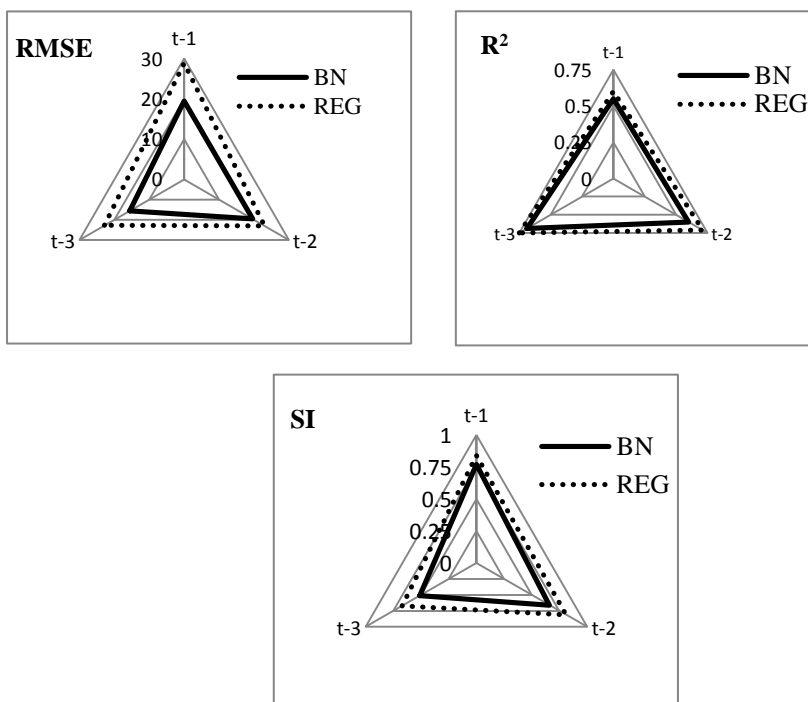
جدول ۴- نتایج رگرسیون خطی چند متغیره در مراحل آموزش و صحت‌سنجی

Table 4. Results multivariate linear regression. used in the training and validation

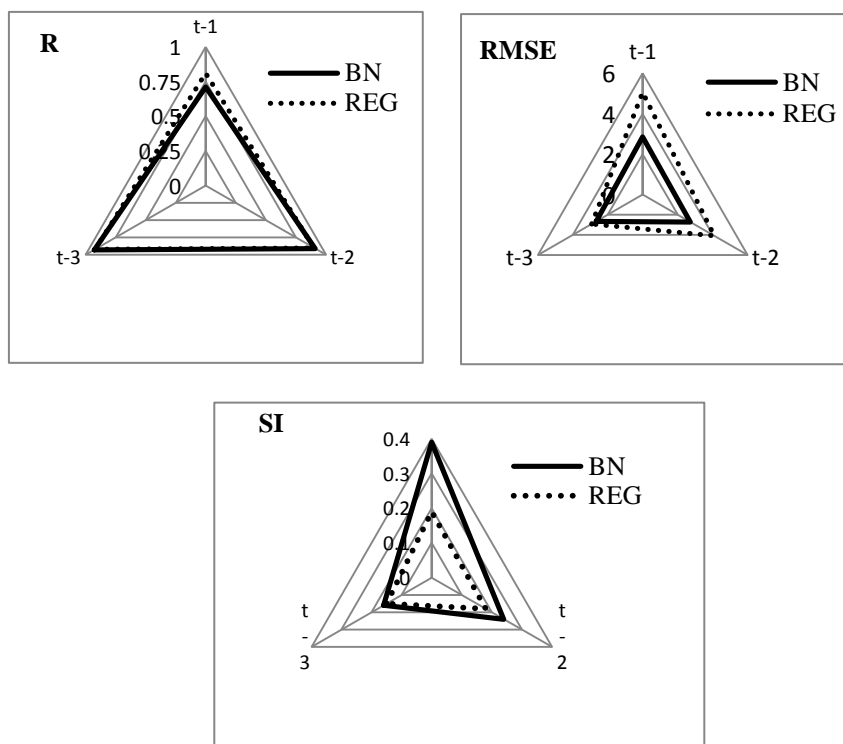
تست			آموزش			پارامترهای ورودی	ترکیب	پارامتر
RMSE(mm, °C, MJ/m ² month)	SI	R ²	RMSE(mm, °C, MJ/m ² month)	SI	R ²			
۵۰/۴۸	-/۸۴	-/۶	۲۸/۸	-/۲	-/۶۳	P _{t-1}	۱	P
۴۰/۰۸	-/۸۱	-/۷۱	۲۳/۰۱	-/۷۱	-/۶۴	P _{t-1} , P _{t-2}	۲	
۲۱/۶	-/۶۷	-/۷۵	۲۲/۸۳	-/۶۵	-/۷۰	P _{t-1} , P _{t-2} , P _{t-3}	۳	
۵/۰۵	-/۱۹	-/۸۱	۵/۹۵	-/۳۵	-/۷۱	T _{t-1}	۱	T
۴/۰۹	-/۱۸	-/۹۰	۳/۸۳	-/۱۹	-/۹۱	T _{t-1} , T _{t-2}	۲	
۲/۹۲	-/۱۵	-/۹۲	۲/۶۶	-/۱۶	-/۹۳	T _{t-1} , T _{t-2} , T _{t-3}	۳	
۴۰/۴۸	-/۳۵	-/۶۰	۶۹/۹۸	-/۱۹	-/۶۱	S _{t-1}	۱	S
۲۵/۹۴	-/۳۲	-/۷۰	۵۱/۱۷	-/۱۷	-/۶۶	S _{t-1} , S _{t-2}	۲	
۲۰/۷۲	-/۲۰	-/۷۸	۵۰/۷۶	-/۱۵	-/۷۵	S _{t-1} , S _{t-2} , S _{t-3}	۳	

(R^2) در ترکیب سوم بیشترین مقدار را به خود اختصاص داده است. همچنین در مورد شاخص پراکندگی (SI) و جذر میانگین مربعات خطا (RMSE) ترکیب سوم کمترین مقدار را دارا می‌باشد. نتایج حاصل از شبیه‌سازی با رگرسیون خطی چندمتغیره نشان می‌دهد میزان ضریب تبیین (R^2) در بخش آموزش و صحت سنجی در بارش، دما و تابش در ترکیب سوم بیشترین مقدار را به خود اختصاص داده است. همچنین در مورد شاخص پراکندگی (SI) و جذر میانگین مربعات خطا (RMSE) ترکیب سوم بهترین حالت را دارد.

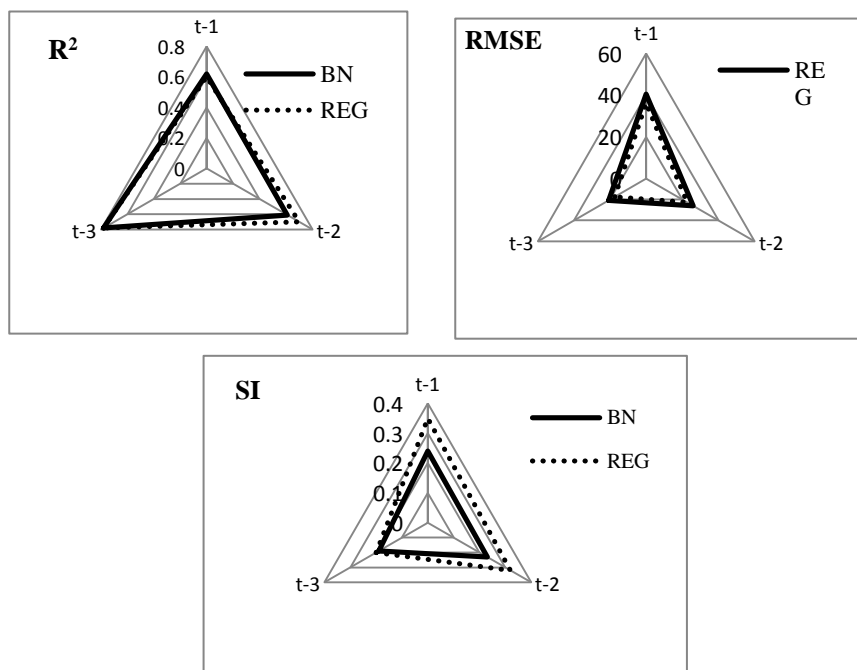
با توجه به جداول ۳ و ۴ ملاحظه می‌گردد در هر سه مورد بارش، دما و تابش ترکیب سوم یعنی در حالتی که ورودی مدل تا سه تأخیر زمانی مورد استفاده قرار می‌گیرد نتایج بهتری ارائه می‌دهد. در مقاله رمضان و همکاران (۱۹) از روش آنتروپی جهت تعیین ورودی‌های مدل شبکه عصبی مصنوعی استفاده شد که نتایج نشان داد استفاده از آنتروپی باعث بهبود شبکه عصبی مصنوعی شد و نتایج بهتری ارائه داد. در پژوهش حاضر، در مدل شبکه بیزین برای بارش، دما و تابش در بخش آموزش و صحت سنجی میزان ضریب تبیین



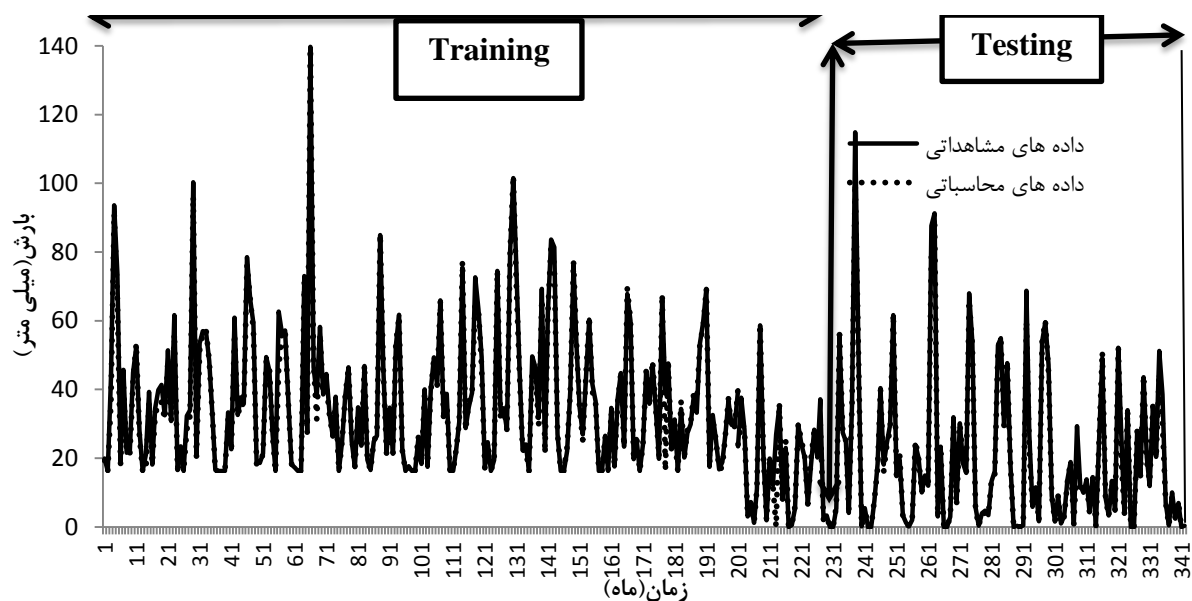
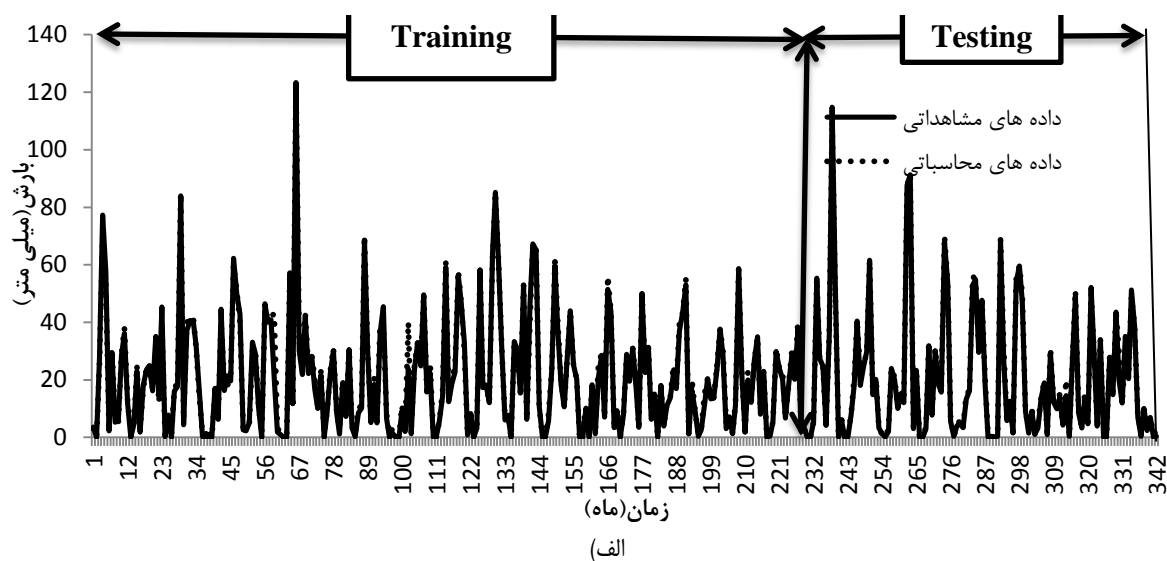
شکل ۴- عملکرد مدل شبکه بیزین و رگرسیون چند متغیره در مرحله تست برای پیش‌بینی بارش ایستگاه سینوپتیک تبریز
 Figure 4. The performance of Bayesian network and multivariate linear regression in the testing for predicted precipitation in Tabriz synoptic station



شکل ۵- عملکرد مدل شبکه بیزین و رگرسیون چند متغیره در مرحله تست برای پیش‌بینی دما ایستگاه سینوپتیک تبریز
 Figure 5. The performance of Bayesian network and multivariate linear regression in the testing for predicted temperature in Tabriz synoptic station

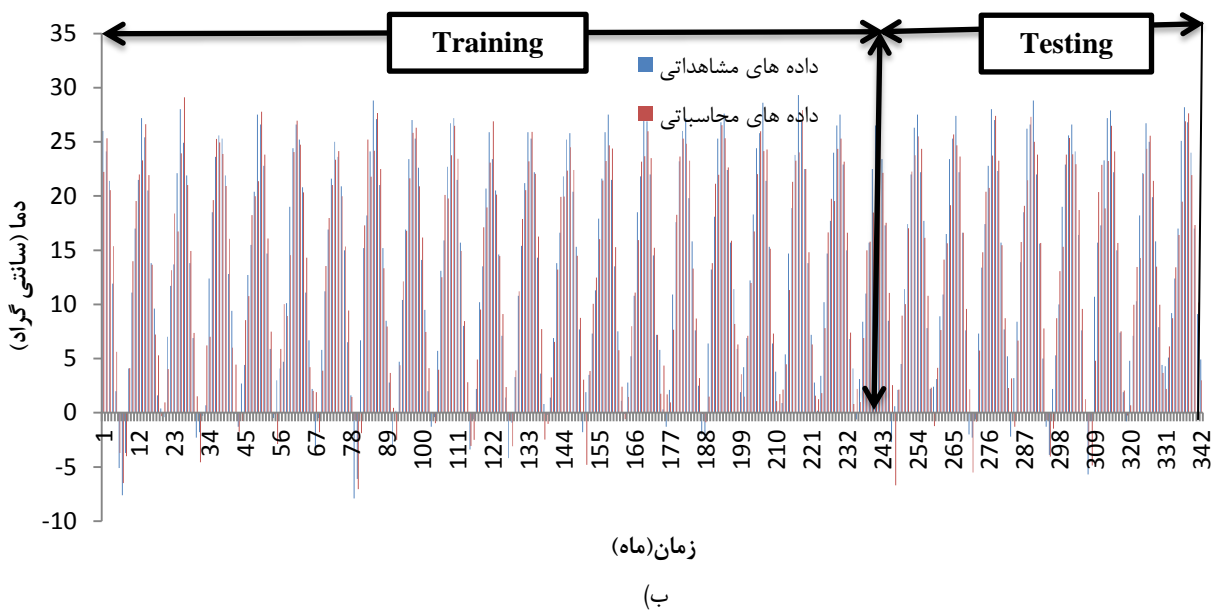
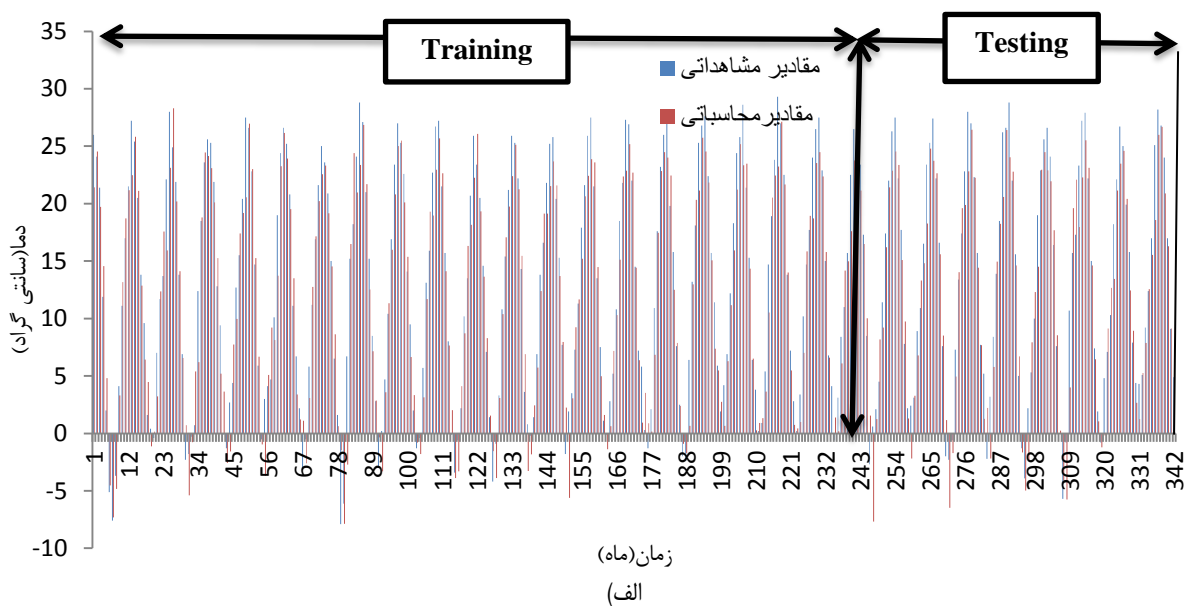


شکل ۶- عملکرد مدل شبکه بیزین و رگرسیون چند متغیره در مرحله تست برای پیش‌بینی تابش ایستگاه سینوپتیک تبریز
 Figure 6. The performance of Bayesian network and multivariate linear regression in the testing for predicted radiation in Tabriz synoptic station



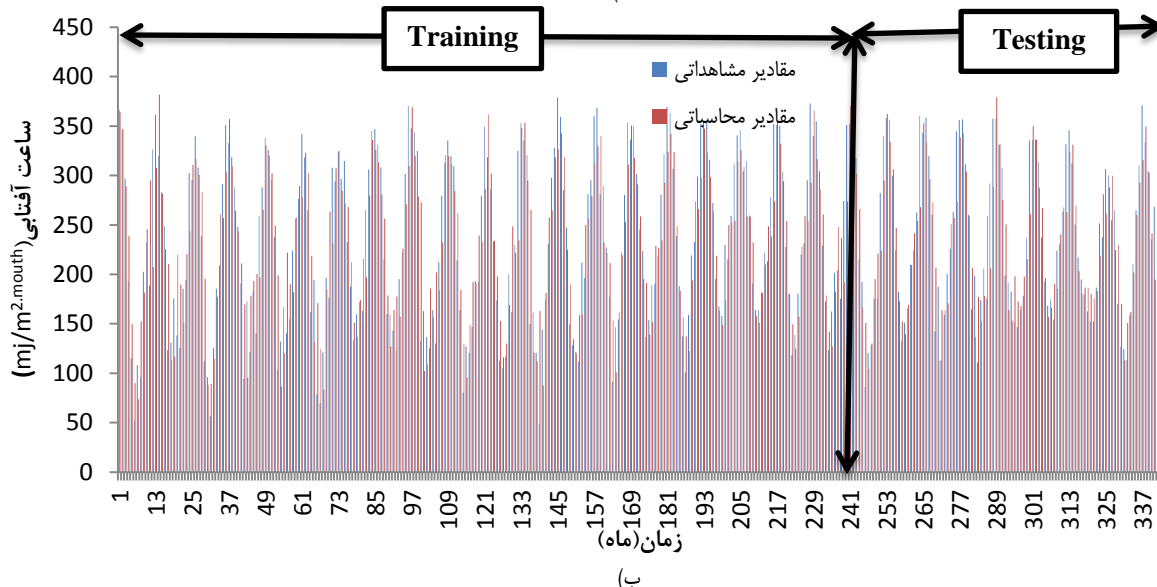
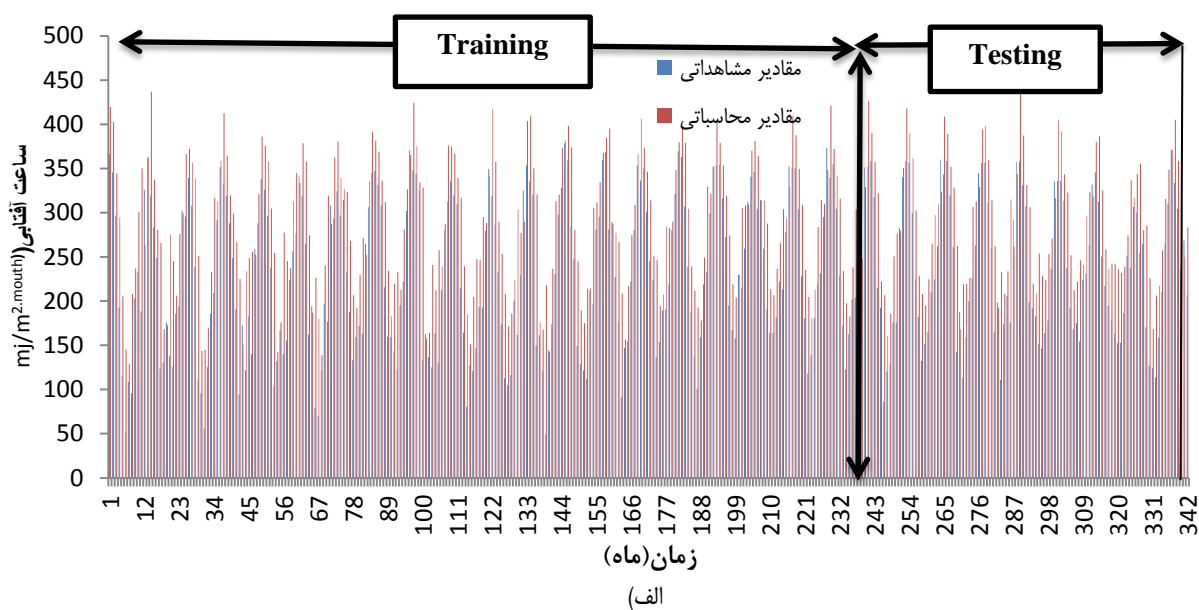
شکل ۷- الف و ب) مقادیر مشاهداتی و پیش‌بینی شده بارش ایستگاه سینوپتیک تبریز با شبکه بیزین و رگرسیون خطی چندمتغیره (ترکیب ۳) بخش آموزش و صحت سنجی

Figure 7. The observed and predicted values of precipitation from Tabriz synoptic station with Bayesian network and multivariate linear regression (compound3) the training and validation



شکل ۸- الف و ب) مقادیر مشاهده‌ای و پیش‌بینی شده دمای ایستگاه سینوپتیک تبریز با شبکه بیزین و رگرسیون خطی چندمتغیره (ترکیب ۳) بخش آموزش و صحت سنجی

Figure 8. The observed and predicted values of temperature from Tabriz synoptic station with Bayesian network and multivariate linear regression (compound3) the training and Testing



شکل ۹- الف و ب)- مقادیر مشاهده‌ای و پیش‌بینی شده ساعات آفتابی ایستگاه سینوپتیک تبریز با شبکه بیزین و رگرسیون خطی چندمتغیره (ترکیب ۳) بخش آموزش و صحت سنجی

Figure 9. The observed and predicted values of radiation from Tabriz synoptic station with Bayesian network and multivariate linear regression (compound3) the training and validation

تحقیق رمضان و همکاران (۱۹) که اذعان داشتند تئوری آنتروپی شانون برای تعیین ورودی مدل‌های هوشمند کارایی دارد مطابقت دارد. با توجه به نتایج به‌دست‌آمده در بخش مدل‌سازی، می‌توان شبکه بیزین را به‌عنوان مدلی با دقت و سرعت پردازش بالا در مدل‌سازی مسائل هیدرولوژیکی و مسائلی با پارامترهای ورودی تأثیرگذار بر پارامتر خروجی، پیشنهاد نمود. همچنین با توجه به نتایج به‌دست‌آمده، تئوری آنتروپی را به‌عنوان روش دقیق و بهینه در جهت کاهش مراحل سعی و خطا در تعیین پارامترهای ورودی در مدل‌سازی استفاده کرد.

در این تحقیق سعی بر آن شد که عملکرد تئوری آنتروپی شانون برای تعیین ترکیب بهینه ورودی‌های مدل‌های هوشمند در سری‌های زمانی موردبررسی قرار گیرد. مقادیر ورودی مشاهده‌ای بارش، دما و تابش در مدل‌های شبکه بیزین و رگرسیون خطی چند متغیره با استفاده از معیارهای ارزیابی موردبررسی قرار گرفت. با استفاده از این معیارها می‌توان گفت تئوری آنتروپی شانون برای تعیین ورودی‌های مدل‌های هوشمند می‌تواند مورد استفاده قرار گیرد. همچنین با استفاده از معیارهای ارزیابی می‌توان گفت شبکه بیزین که در آن ورودی مدل از ترکیب پارامترها با سه تأخیر زمانی تشکیل شده، از دقت قابل قبولی برخوردار است. این نتایج با

منابع

1. Amorocho, J. and B. Espildora. 1973. Entropy in the assessment of uncertainty in hydrologic systems and models. *Journal of Water Resource. Research*, 9(6): 1551-1522.
2. Al-Zahrani, M. and T. Husain. 1998. An algorithm for designing a precipitation network in the south-western region of Saudi Arabia. *Journal of Hydrology*, 205: 205-216.
3. Botsis, D., P. Latinopoulos, K. Diamantaras. 2012. Investigation of The Effect of Interception and Evapotranspiration on the rain fall-Run off Relationship using Bayesian Networks, 12th International Conference on Environmental Science and Technology (CEST), 8 - 10 September, Rhodes, Greece
4. Brunsell, N.A. 2010. A multistate information theory approach to assess spatial-temporal variability of daily precipitation. *Journal of Hydrology*, 385: 165-172.
5. Carmona, G., J.L. Molina, J. Bromley, C. Varela-Ortega and J.L. Garcia-Arostegu. 2011. Object Oriented Bayesian network for participatory water management, two case Studies in Spain, *Journal of Water resources planning and management*, 137: 366-376.
6. Chiang, W. and Y. Hui-Chung. 2014. Spatiotemporal Scaling Effect on Rainfall Network Design Using Entropy. *Journal of Entropy in Hydrology*, 16: 4626-4647.
7. Chen, Sh. 2015. Mining Informative Hydrologic Data by Using Support Vector Machines and Elucidating Mined Data according to Information Entropy, *Journal of entropy*, 17: 1023-1041.
8. Farmani, R., H.J. Henriksen and D. Savic. 2009. An evolutionary Bayesian belief network methodology for optimum management of groundwater contamination, *Journal of Environmental Modeling & Software*, 24: 303-310.
9. Farajzadeh, J., A. FakheriFard and S. Lotfi. 2014. Modeling of monthly rainfall and runoff of Uremia lake basin using feed-forward neural network and time series analysis model. *Journal of Water Resources and Industry*, 7(8): 38-48.
10. Harmancioglu, N.B. and N. Alpaslan. 1992. Water quality monitoring network design: A problem of multi-objective decision making. *Journal of Water Resource. Bull*, 28(1): 179-192.
11. Harmancioglu, N.B. 1984. Entropy concept as used in determination of optimum sampling intervals. *Proc. of Hydrosoft 84, International Conference on Hydraulic Engineering Software*, September 10-14, 1984. Protozoa, Yugoslavia, pp: 6-99 and 6-110.
12. Jha, R. and V.P. Singh. 2008. Evaluation of river water quality by entropy. *Journal of KSCE Civil Engineer*, 12(1): 61-69.
13. Karamouz, M., A.K. Nokhandan, R. Kerachian and C. Maksimovic. 2009. Design of on-line river water quality monitoring systems using the entropy theory: a case study. *Journal of Environmental modeling and Assessment*, 155(1-4): 63-81.
14. Karimi Hoesini, A. 2009. Compare the methods of locating the rain-gauge stations in the GIS environment. Master's thesis, Faculty of Agriculture, Department of Irrigation and Reclamation, agricultural meteorology, Tehran University, Supervisor: Abdul Horfar, great hope Haddad (In Persian).
15. Misra, D., T. Oommen. A. Agarwal and S.K. Mishra. 2009. Application and analysis of Support Vector machine based simulation for runoff and sediment yield, *Journal of Bio Systems Engineering*, 103: 527-535.
16. Mishra, A.K. and P. Coulibaly. 2010. Hydrometric network evaluation for Canadian watersheds. *Journal of Hydrology*, 380: 420-437.
17. Masoumi, F. and R. Kerachian. 2008. Underground water quality assessment and performance monitoring systems with discrete entropy, The second national conference, Esfahan (In Persian).
18. Montesarchio, V. and F. Napolitano. 2010. A single-site rainfall disaggregation model based on entropy. *International Workshop Advances in Statistical Hydrology*. May 23-25, Taormina, Italy.
19. Remesan, R., A. Ahmadi, M.A. Shamim and D. Han. 2010. Effect of data time interval on real-time flood forecasting, *Journal of hydrology, informatics*, 12(4): 396-407.
20. Shannon, E.A. 1948. *Mathematical Theory of Communication*. Bell System Technical Journal 27: 379-423.
21. Singh, V.P. and K. Singh. 1985. Derivation of the Pearson type (PT)-III distribution by using the principle of maximum entropy (POME), *Journal of Hydrology*, 80: 197-214.
22. Singh, V.P. and D.A. Woolhiser. 2002. Mathematical modeling of watershed hydrology. *Journal of Hydrologic engineering*, 7(4): 270-292.
23. Sonuga, J.O. 1972. Principle of maximum entropy in hydrology frequency analysis. *J. Hydrol*, 17: 177-191.
24. Yue, L.P., Q. Hui and W. Jian-Hu. 2010. Groundwater quality assessment based on improved water quality index in Pengyang Country, Ningxia, northwest China, *Journal of Chem*, 7(S1): 209-216.

Applying Shannon Entropy in Bayesian Network Input Preprocessing For Time Series Modeling

Fateme Akhoni Pourhosseini¹, Mohammad Ali Gorbani² and Kaka Shahedi³

1- M.Sc. Student of Water Resources Engineering, University of Tabriz
(Corresponding author: fateme.pourhosseini@yahoo.com)

2- Associate Professor, of Water Resources Engineering, University of Tabriz

3- Associate Professor, Faculty of Natural Resources, Sari Agricultural Sciences and Natural Resources, University

Received: May 31, 2016

Accepted: January 25, 2017

Abstract

Selecting appropriate inputs for intelligent models is important due to reduce costs and save time and increase accuracy and efficiency of models. The purpose of this study is using Shannon entropy to select the optimum combination of input variables in time series modeling. Monthly time series of precipitation, temperature and radiation in the period of 1982-2010 was used from Tabriz synoptic station. Precipitation, temperature and radiation parameters with different delays are considered as input to the Shannon entropy. The results showed that time series with three delays provide the better results for the modeling. Applying Bayesian network and multivariate linear regression analysis were performed. Models performance was evaluated using three criteria: coefficient of determination (R^2), root mean square error (RMSE), and the dispersion. Index (SI). The results indicated that Bayesian neural network model shows the best performance to simulate time series of precipitation, temperature and radiation in compare to multivariate linear regression analysis. The results showed that Shannon entropy has better performance in selection of the appropriate entry into intelligent models.

Keywords: Entropy, Multivariate linear regression, Time series, Bayesian network